Estimating Peaks of Stationary Random Processes: A Peaks-over-Threshold Approach

Dat Duthinh, M.ASCE[1]; Adam L. Pintar[2]; and Emil Simiu, F.ASCE[3]

Abstract: Estimating properties of the distribution of the peak of a stochastic process from a single finite realization is a problem that arises in a variety of science and engineering applications. Furthermore, it is often the case that the realization is of length T whereas the distribution of the peak is sought for a different length of time, $T_1 > T$. The procedure proposed here is based on a peaks-over-threshold extreme value model, which has an advantage over classical models used in epochal procedures because it often results in an increased size of the relevant extreme value data set. For further comparison, the translation approach depends upon the estimate of the marginal distribution of a non-Gaussian time series, which is typically difficult to perform reliably. The proposed procedure is based on a two-dimensional Poisson process model for the pressure coefficients y, above the threshold B. The estimated distribution of the peak value depends upon the choice of the threshold. The threshold choice is automated by selecting the threshold that minimizes a metric that captures the trade-off between bias and variance in estimation. Two versions of the proposed new procedure are developed. One version, denoted FpotMax, includes estimation of a tail length parameter with a similar interpretation of the generalized extreme value distribution tail length parameter. The second version, denoted GpotMax, assumes that the tail length parameter vanishes. **DOI: 10.1061/AJRUA6.0000933.**

Author keywords: Autocorrelation; Bootstrap; Decluster; Gumbel distribution; Independent peaks; Monte Carlo simulation; Peaks-over-threshold; Poisson process; Stationary time series; Wind pressure.

Introduction

Estimating properties of the distribution of the peak of a stochastic process from a single finite realization is a problem that arises in a variety of science and engineering applications. Furthermore, it is often the case that the realization is of length T whereas the distribution of the peak is sought for a different length of time, $T_1 > T$. An example of interest in structural engineering, which is the focus of this paper, is the time history of pressure coefficients measured on a model in an aerodynamics laboratory.

Cook and Mayne (1979, 1980) proposed a simple procedure in which the time history of the pressure coefficients of length T is divided into n equal segments. A sample is created consisting of the peaks of each of those segments, and a Gumbel cumulative distribution function (CDF) is fitted to that sample. To obtain the largest peak for a time history of length $T_1 = rT/n\,(r \geq n)$, that CDF is raised to the rth power, resulting in a second Gumbel distribution describing the peak of a series of length T_1. This procedure is commonly implemented using the best linear unbiased estimate (BLUE) (Lieblein 1974) for the parameters of the Gumbel distribution. This method is henceforth referred to as the epochal method.

Several other methods have been proposed for this problem. For the particular case in which the marginal distribution of the process is Gaussian, closed-form expressions for the mean and standard deviation of the distribution of the peak are available (Rice 1954; Davenport 1964). If the distribution is not Gaussian, Sadek and Simiu (2002) developed a nonlinear mapping procedure, sometimes referred to in the literature as translation, by which those statistics can be obtained. The translation method depends heavily on the engineer's ability to choose an appropriate marginal probability distribution. In practice, because of the difficulty of this task, the performance of the translation method can be unsatisfactory.

Another general approach for winds is the so-called method of independent storms described and improved upon over several papers, e.g., Cook (1982), Peterka (1983), Harris (1999), and Harris (2009). This approach argues for finding and using in estimation, the single largest value in each available independent storm. The approach does not leverage asymptotically justifiable models such as the generalized extreme value distribution, the generalized Pareto distribution, or the two-dimensional Poisson process model that provides the bedrock of the proposed method. In fact, Harris (2009) argues against the applicability of methods based on models derived from asymptotic theory. However, although in practice infinite sample sizes are never achieved, these models, especially the two-dimensional Poisson process model in this paper, remain popular and seemingly useful, see Smith (1989), Smith (2004), Coles (2004), and Mannshardt-Shamseldin et al. (2010) for example.

This paper (1) describes a procedure for estimating the distribution of the peak based on a peaks-over-threshold (POT) model for extreme values; and (2) assesses its performance through cross-validation. The authors do not claim to undertake a detailed comparison of their method with all other alternatives. Two major contributions of this procedure to the existing literature are (1) estimation of the distribution of the peak values using a POT model for which there is no analytical solution because the number of

[1]Research Structural Engineer, Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899 (corresponding author). E-mail: dduthinh@nist.gov

[2]Mathematical Statistician, Statistical Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD 20899. E-mail: adam.pintar@nist.gov

[3]NIST Fellow, Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899. E-mail: simiu@nist.gov

Note. This manuscript was submitted on April 18, 2016; approved on May 26, 2017; published online on September 11, 2017. Discussion period open until February 11, 2018; separate discussions must be submitted for individual papers. This paper is part of the ***ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering***, © ASCE, ISSN 2376-7642.

threshold crossings is itself a random quantity, and (2) a rigorous approach to quantifying uncertainty for the entire estimated distribution of the peak, not just, for example, the mean of that distribution. To carry out the calculations associated with this approach, software that leverages the *R* environment for statistical computing and graphics is available at Pintar (2016). Instructions for installation and use are available there as well.

Peak Estimation by a Peaks-over-Threshold Model

Peaks-over-threshold models are applied to observations, y, that exceed a specified threshold, B, of the time series being considered. All peaks over a high threshold are used, instead of epochal peaks for epochs of fixed size, because this choice will generally allow the use of more observations from the original time series to infer the values of the model parameters, potentially leading to less uncertainty.

Although the results of the POT approach depend upon the choice of threshold, a procedure for objectively selecting an appropriate threshold using the data themselves eliminates the dependence of the final result on an arbitrary choice. The threshold selection procedure is summarized subsequently and described fully by Pintar et al. (2015).

The complete procedure is described and illustrated in what follows with reference to data set jp1 in the NIST–University of Western Ontario (UWO) Aerodynamic Database for Rigid Buildings (NIST 2004; Ho et al. 2005), Building 7, in open terrain. The building was modeled at a scale of 1:100, tested for wind directions ranging from 0° to 90° and from 270° to 360° every 5°, and data were collected for 100 s at 500 Hz. Because of building symmetry, for terrain with the same exposure in all directions, only half of all possible wind directions need to be investigated. In Fig. 1, 0° and 90° are in the $+y$ and the $+x$ directions, respectively. Normalized pressure coefficients (Fig. 2) from Tap 1715, in the middle of the roof edge, were investigated for wind direction 270°. The procedure is described in the following subsections.

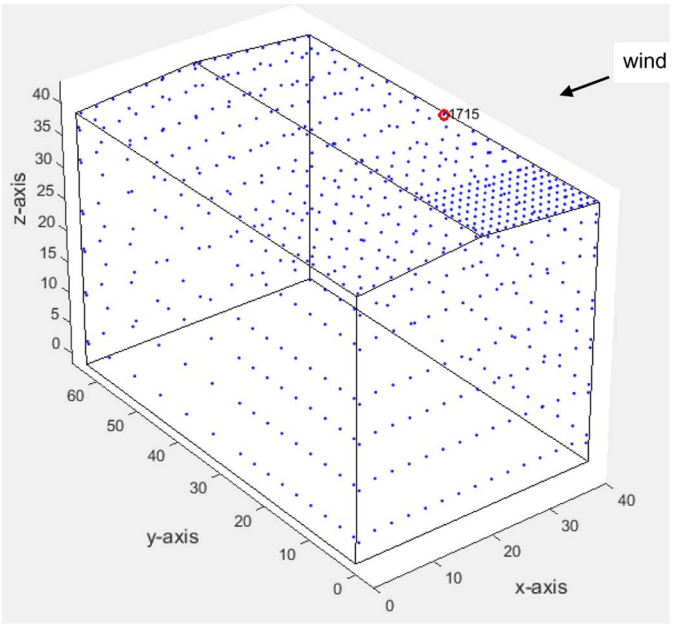

Fig. 1. Building 7, data set jp1 (NIST–UWO database); dimensions in feet (1 ft = 0.3048 m); roof slope 1:12

Step 1: Reverse Signs the Time Series If Necessary

The procedure is developed for maxima. Because the peaks of interest in Fig. 2 are minima, the signs of this time series were reversed. If analysts are interested in both peaks and valleys, they should apply the method twice, first with the original signs and a second time with reversed signs.

Step 2: Choose a Model

The POT model leveraged in this work is a two-dimensional Poisson processes defined by the intensity function λ in either Eq. (1) or Eq. (2). Note that Eq. (2) is derived by taking the limit of Eq. (1) as $k(t)$ approaches zero. The appropriateness of such models for crossings of a high threshold, B, was first discussed by Pickands (1971), and the form of Eq. (1) is given, for example, in Eq. (1.19) of Smith (2004)

$$\lambda(t,y) = \frac{1}{\sigma(t)}\left[1 + \frac{[y-\mu(t)]}{\sigma(t)}\right]_+^{-1-1/k(t)} \tag{1}$$

$$\lambda(t,y) = \frac{1}{\sigma(t)}\exp\left\{\frac{-[y-\mu(t)]}{\sigma(t)}\right\} \tag{2}$$

Note that the threshold, B, does not appear in either Eq. (1) or Eq. (2). It instead specifies the domain on which the Poisson process is defined. Specifically, if data are collected over the time interval $[t_1, t_2]$, the domain of the Poisson process is $[t_1, t_2] \times [B, \infty)$. It is possible to allow the threshold to vary with time t, which is indeed the case in Pintar et al. (2015), but the dependence on t is suppressed here to be able to describe the domain more succinctly. General Poisson processes are described in many texts, for example, Chapter 4 of Resnick (1992). Briefly, for the two-dimensional case, if D is some two-dimensional region, the number of observations found in D follows a Poisson distribution with mean equal to the volume trapped by the intensity function over D. Furthermore, if D_1 and D_2 are disjoint regions, the numbers of observations falling in D_1 and D_2 are independent of each other. The $+$ subscript in Eq. (1) means that negative values inside the square brackets are raised to zero. Of particular interest, and the focus of the wind tunnel examples in this paper, is the case when the parameters are constant, i.e., $\mu(t) = \mu$, $\sigma(t) = \sigma$, and $k(t) = k$, although Pintar et al. (2015) consider time-varying parameters to capture the behavior of a mixed wind climate outdoors. The parameters μ and σ are the location and scale parameters, respectively. Eq. (2) is the limit of Eq. (1) as the tail length parameter k approaches zero, just as the Gumbel distribution is the limit of the generalized extreme value distribution as the tail length parameter limits to zero. For this reason, the complete approach based on the Poisson process defined by the intensity function in Eq. (2) is henceforth designated as

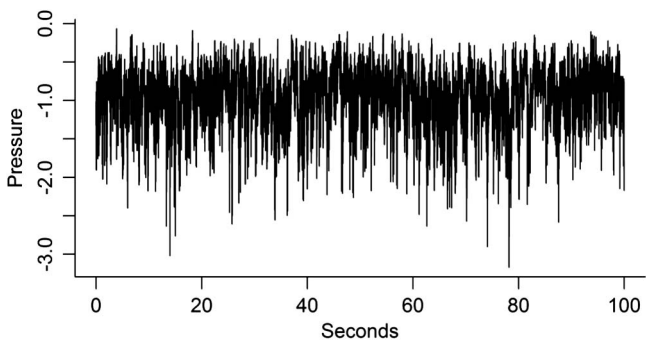

Fig. 2. Pressure coefficients from Tap 1715 for wind in $-x$ direction

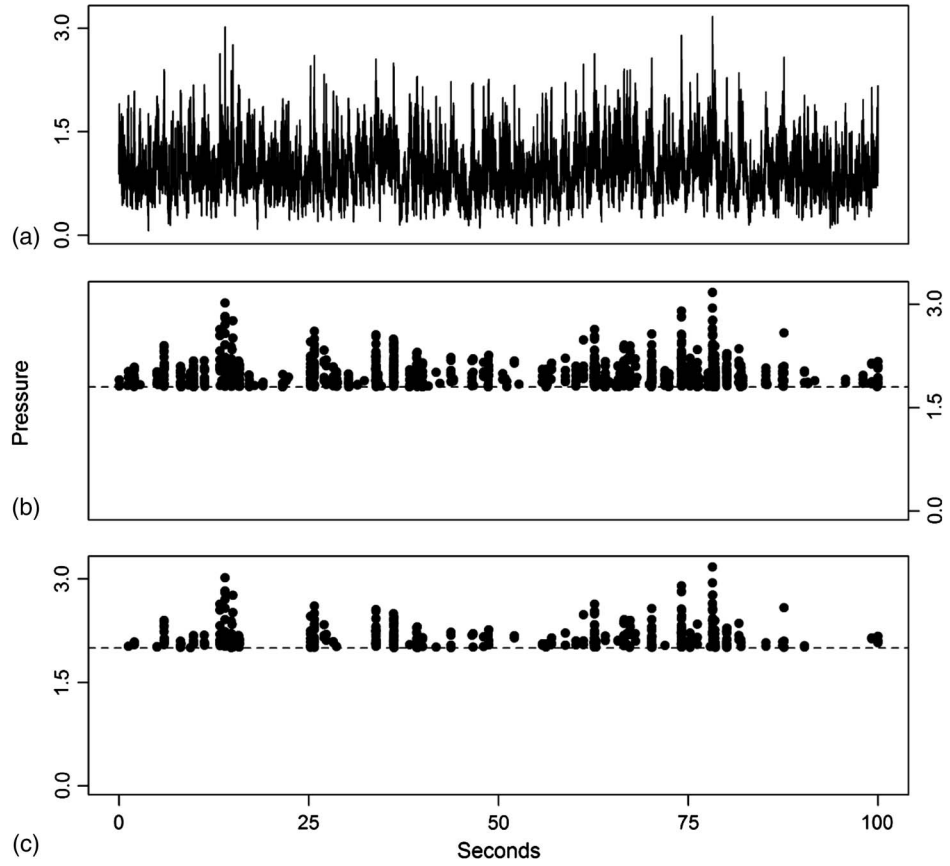

Fig. 3. (a) Raw time series; (b) observations above 1.8; (c) observations above 2.0

GpotMax, whereas the Poisson process with $\lambda(t, y)$ defined in Eq. (1) is designated as the full model, or FpotMax.

Step 3: Decluster

Fig. 3 depicts the same raw time series as Fig. 2 [Fig. 3(a) reverses the signs of Fig. 2], and two thresholded variants with the thresholds $B = 1.8$ [Fig. 3(b)] and $B = 2.0$ [Fig. 3(c)]. Note in Fig. 3 that successive peaks can be separated by time intervals smaller than the time between an up-crossing of the mean and the subsequent down-crossing of the mean. Such successive peaks are typically strongly correlated, as shown by Fig. 4, which plots the values of the history in Fig. 3(a) on the horizontal axis and their immediately adjacent neighbors on the vertical axis. Poisson processes are not appropriate for histories in which immediately adjacent neighbors are highly correlated because of the independence assumption that underlies them.

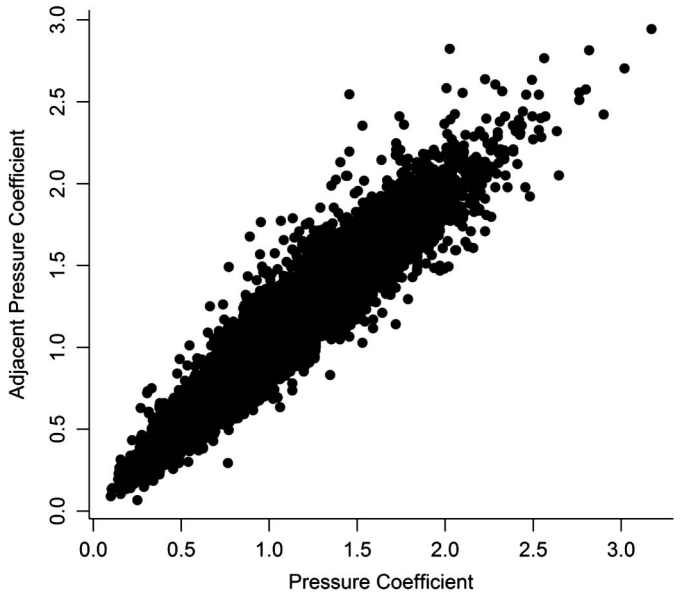

Fig. 4. Scatter plot of adjacent pressure coefficients from Fig. 3(a)

Clusters are data blocks within time intervals defined by an up-crossing of the mean and the subsequent down-crossing of the mean. All but the cluster maximums are discarded. Fig. 5 displays the counterparts of Fig. 3 after declustering, and Fig. 6 depicts the counterpart of Fig. 4. Fig. 6 shows that declustering is indeed very effective. The approach to creating clusters was chosen for its convenience and because it has proved effective in removing the correlation between adjacent observations. If in a particular example a different approach that is natural or physically meaningful is available, it should be used instead. Declustering is intended only to allow adjacent observations to be assumed statistically independent.

Note that similar declustering procedures have been used in past work on extremes. For example, the declustering procedure used here is similar to that employed by Smith (1989) for ozone extremes. Furthermore, the approach to identifying the "lulls" of Cook (1982) also bears resemblance. A distinction between the approach used here and that of Cook (1982) is Cook's absolute velocity of 5 m/s that identifies the start of a lull. It is presumed that the choice of 5 m/s is physically meaningful, whereas the corresponding choice in this paper of the mean of the history under study is convenient yet sensible, and has proven to work well in practice as demonstrated in Figs. 4 and 6.

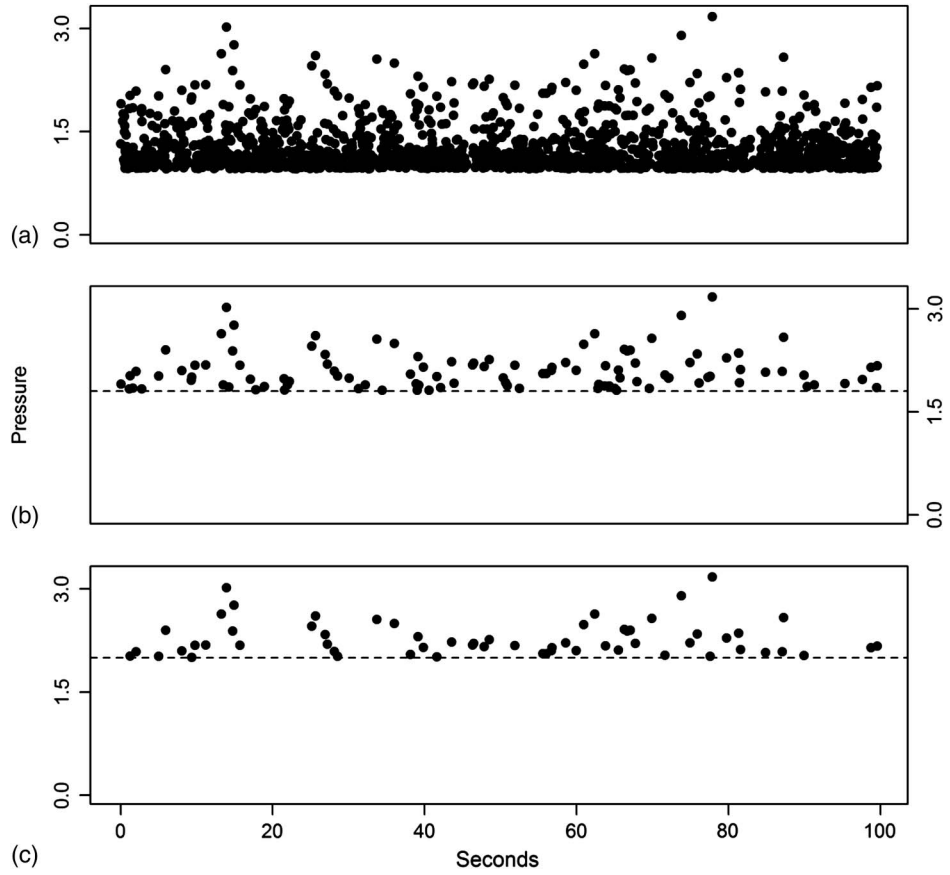

Fig. 5. (a) Declustered time series; (b) resulting observations above 1.8; (c) above 2.0

Step 4: Select Optimal Threshold

Historically, a hurdle to the use of the POT models has been the appropriate choice of a threshold. Because the threshold dictates the data that are included in (or omitted from) the sample used to fit the model, its impact on the results can be large. The model becomes more appropriate as the threshold increases, but the threshold cannot be too high because too few observations will remain for fitting the model since observations are taken over only a finite period of time. Any approach to choosing a threshold must balance these competing aspects. A common and easy to implement—although not necessarily optimal—approach is to pick a high quantile of the series, e.g., 95% (Mannshardt-Shamseldin et al. 2010, p. 489). This paper takes the approach of Pintar et al. (2015). An optimal threshold based on the fit of the model to the data, as judged by the W statistics defined in Eq. (1.30) of Smith (2004), is used

$$W = \frac{1}{k(t)} \log\left\{1 + \frac{k(t)y}{\sigma(t) + k(t)[B(t) - \mu(t)]}\right\} \tag{3}$$

The limit of Eq. (3) as $k(t)$ approaches zero is

$$W = \frac{y}{\sigma(t)} \tag{4}$$

Selection of the threshold begins by specifying a sequence of thresholds to examine. The sequence is defined by the minimum and maximum number of observations to be used in estimation. As a minimum number of observations, 10 and 15 have been found to work reasonably well for Eqs. (2) and (1), respectively, providing five observations per parameter. The maximum number of observations chosen should be sufficiently high so that it will exceed the number of observations corresponding to the optimal threshold and yet still be feasible from a computational perspective. For each threshold in the sequence, the model parameters are estimated by maximum likelihood, which is discussed next in Step 5. The W statistics are then calculated by plugging the estimated

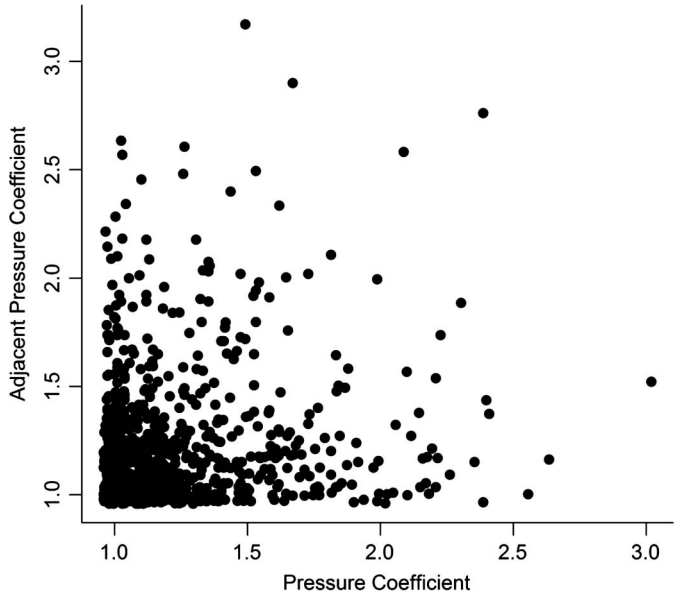

Fig. 6. Scatter plot of adjacent pressure coefficients from Fig. 5(a)

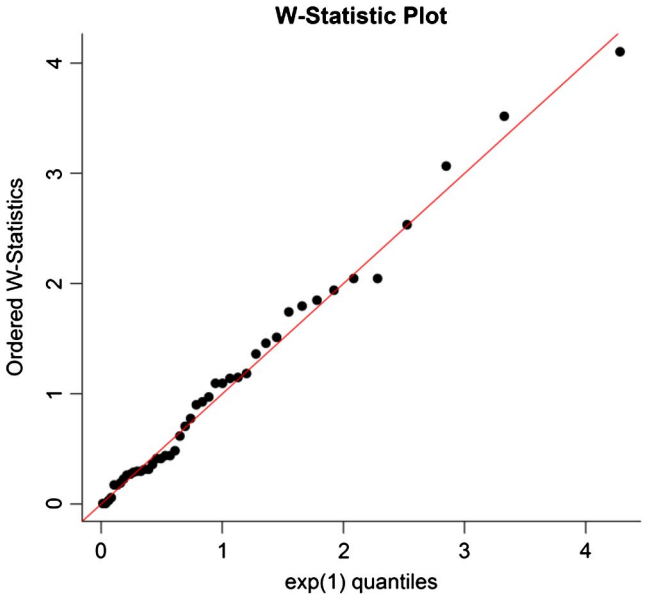

Fig. 7. Plot of the W statistics for the zero-tail model (GpotMax) versus their corresponding standard exponential quantiles for the declustered series depicted in Fig. 5(a) using the optimal threshold (threshold = 2.10, 45 data points)

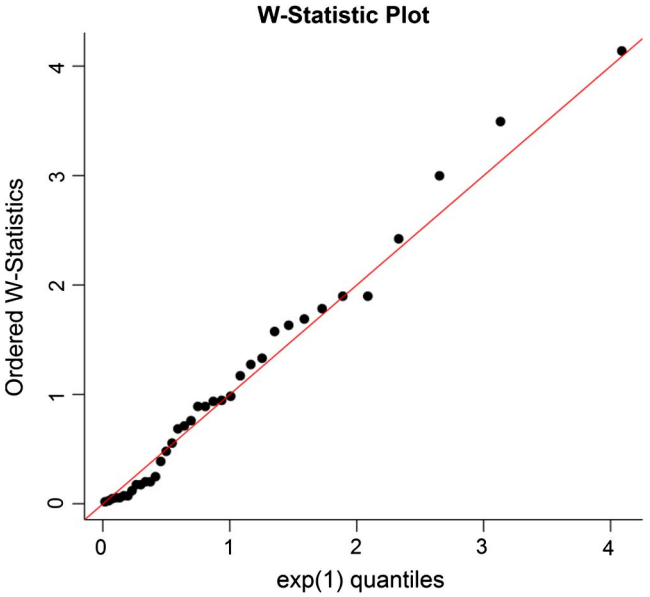

Fig. 8. Plot of the W statistics for the full model (FpotMax) versus their corresponding standard exponential quantiles for the declustered series depicted in Fig. 5(a) using the optimal threshold (threshold = 2.16, 37 data points)

parameters into either Eq. (3) or Eq. (4), and they are then plotted against the quantiles of the standard exponential distribution, which is the probability distribution of W if the data are perfectly described by the Poisson process model being used. For a perfect fit, the points would fall exactly on the diagonal line. Each plot, one for each threshold in the sequence of thresholds, is summarized by the maximum vertical distance from a point to the 45° line. The threshold that minimizes that distance is selected. This approach to selecting a threshold is also comparable to that used by Pickands (1994). Figs. 7 and 8 show plots of the ordered W statistics versus exponential quantiles for the zero-tail (GpotMax) and full (FpotMax) models, respectively, using the optimal thresholds for the declustered series in Fig. 5(a).

Step 5: Estimate Model Parameters

As previously mentioned, it was necessary to estimate the model parameters, $\eta = (\mu, \sigma, k)$ for the intensity function in Eq. (1) and $\eta = (\mu, \sigma)$ for the intensity function in Eq. (2), for each threshold in a sequence of thresholds when choosing the optimal threshold. Thus, even though parameter estimation is discussed as a separate Step 5, it is a major part of Step 4. Parameter estimation proceeds by maximum likelihood (Casella and Berger 2002, Section 7.2.2). For the intensity function given in Eq. (1), the natural logarithm of the likelihood function (which is often optimized in lieu of the likelihood function for numerical stability) when $\mu(t) = \mu$, $\sigma(t) = \sigma$, $k(t) = k$, and $B(t) = B$ is

$$l(\mu, \sigma, k) = \left(\frac{-1}{k} - 1\right)\left\{\sum_{i=1}^{n} \ln\left[1 + \frac{k(y_i - \mu)}{\sigma}\right]\right\} - n\ln(\alpha) - T\left[1 + \frac{k(B-\mu)}{\sigma}\right]^{-1/k} \quad (5)$$

where n = number of declustered observations that cross the chosen threshold, B. For the intensity function given in Eq. (2), the natural logarithm of the likelihood function is

$$l(\mu, \sigma) = -n\ln(\sigma) - \frac{\sum_{i=1}^{n} y_i}{\sigma} + \frac{n\mu}{\sigma} - T\exp\left[\frac{-(B-\mu)}{\sigma}\right] \quad (6)$$

When selecting the optimal threshold, it is necessary only to estimate the parameters for each potential threshold. However, to later quantify uncertainty, the Hessian matrix of the natural logarithm of the likelihood function at its maximum is needed. This is the primary reason for listing parameter estimation as a separate Step 5. To save on computation in Step 4, calculation of the Hessian matrix is omitted. After the optimal threshold is chosen, the Hessian matrix of the natural logarithm of the likelihood function at its maximum is calculated only for the selected threshold. Note that it is necessary to maximize the functions, and thus calculate the Hessian matrix at the maximum for Eqs. (5) and (6) numerically.

Step 6: Empirically Build Distribution of Peak by Monte Carlo Simulation

A series of desired length is generated from the fitted nonhomogeneous two-dimensional Poisson process model, and the peak of the generated series is recorded. Simulating a nonhomogeneous two-dimensional Poisson process may be accomplished using Algorithm 9 of Pasupathy (2011), for example. Simulation from the estimated model and subsequent recording of the peak value is repeated n_{mc} times. The recorded peaks form an empirical approximation to the distribution of the peak. A histogram of the simulated peaks over 100 s with $n_{mc} = 1{,}000$ for the example data set is shown in Fig. 9, in which the mean value is marked by the triangle, and is the reported final result.

Step 7: Quantify Uncertainty

Recall that the goal is to estimate the distribution of the peak of the time series under study. Thus the uncertainty in the estimate of the entire distribution of the peak is being quantified, not only, for example, the uncertainty in the mean of that distribution. To accomplish this, a second layer of Monte Carlo sampling is included.

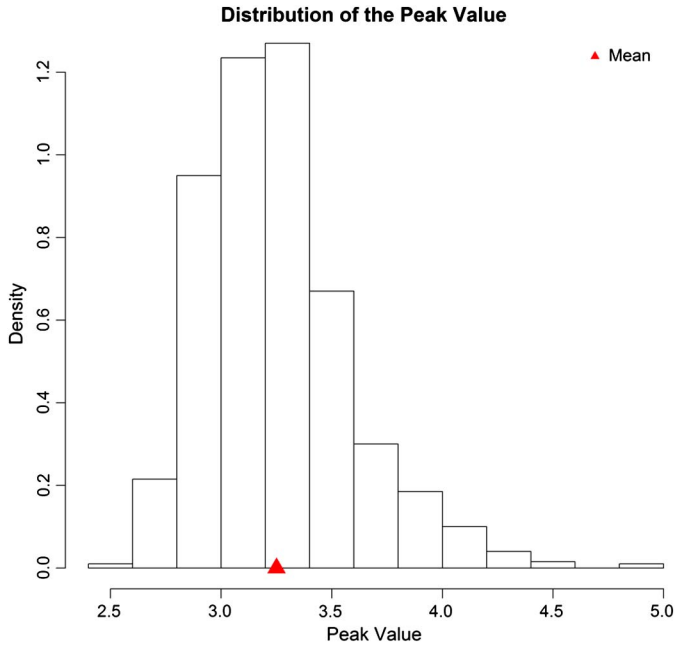

Fig. 9. Histogram of the estimated distribution of the peak value starting with the time series depicted in Fig. 5(a); the triangle shows the mean of the distribution

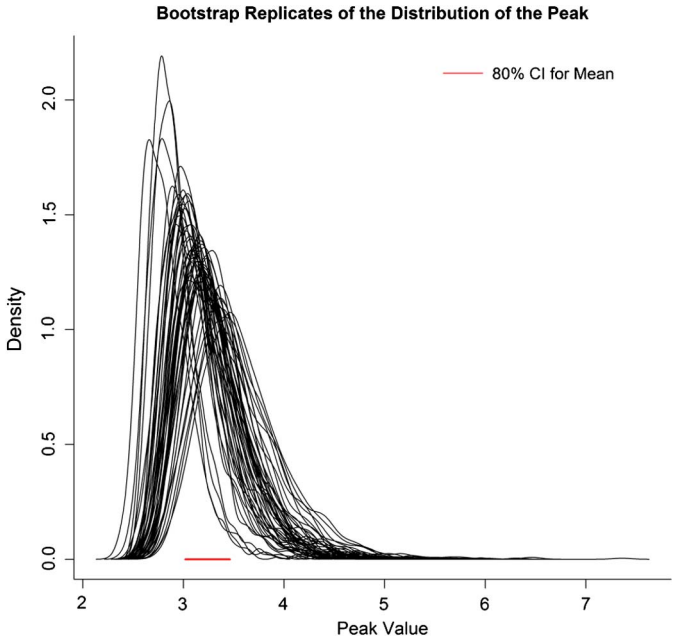

Fig. 10. Bootstrap replicates of the distribution of the peak starting with the series shown in Fig. 5(a); the short horizontal line shows an 80% confidence interval (CI) for the mean of the distribution of the peak value

The input to Step 6 is the maximum likelihood estimate of η, $\hat{\eta}$. However, because only a finite sample is available, these estimates are uncertain. That uncertainty may be described using the multivariate Gaussian distribution and the Hessian matrix calculated in Step 5. More specifically, values of η that are also consistent with the observed time series are sampled, and Step 6 is repeated n_{boot} times for those new parameter values. The mean of the multivariate Gaussian distribution is equal to the estimated parameters, $\hat{\eta}$, and the covariance matrix is equal to the negative inverse Hessian matrix of the natural logarithm of the likelihood evaluated at its maximum. The result of Step 7 is n_{boot} empirical approximations to the distribution of the peak. Fig. 10 shows $n_{\text{boot}} = 50$ replicates of the distribution of the peak for the example data set. Only 50 replicates are shown in Fig. 10 for clarity. Typically, more replicates, perhaps 1,000, would be desired. The bar shown in Fig. 10 depicts an 80% confidence interval for the mean, which was calculated from 1,000 replicates. This technique approximates a bootstrap algorithm (Efron and Tibshirani 1994) because the bootstrap distribution of the estimated parameter values is not constructed by resampling the data, but assumed to be a multivariate Gaussian distribution.

Results for Tap 1715

Fig. 11 shows several results for the example series. The application of GpotMax to the whole 100 s series is depicted by the dashed line, and it is close to the observed peak, shown by the solid line. On the same plot, the squares show the results of six analyses performed on six partitions of the same time series, each of length 100/6 s. The GpotMax estimates closely track the observed peaks (circles) for each of the partitions. For each partition of length 100/6 s, GpotMax may also be used to calculate the mean of the distribution of the peak for a 100-s duration, shown by the triangles in Fig. 11. The six individual partitions can yield estimates that differ by as much as approximately

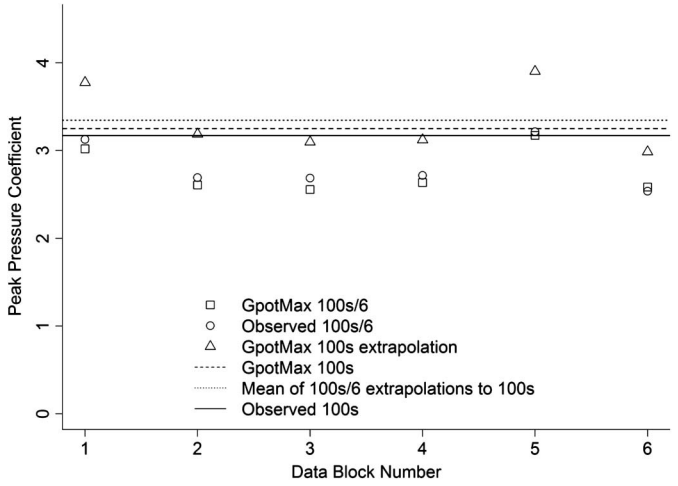

Fig. 11. Comparison of estimates based on six equal data blocks and on global analysis, using GpotMax

25% from the estimate based on the entire 100-s time series. However, the average of these six estimates, shown by the dotted line, is reasonably close to the estimate based on the whole series and the observed 100-s peak.

When FpotMax is applied to the full 100-s series, the estimated mean of the distribution of the peak is 3.17 with a standard error of 0.33. In comparison, for GpotMax, the estimated mean is 3.25 with a standard error of 0.17. That the estimate for FpotMax is smaller than that for GpotMax reflects the slightly negative estimate (−0.02) of the tail length parameter, k. The larger standard error for the FpotMax estimate (almost double in this example) is expected because the number of parameters being estimated increased from two to three. More complex models generally lead to less-certain estimates when the same data are used for estimation.

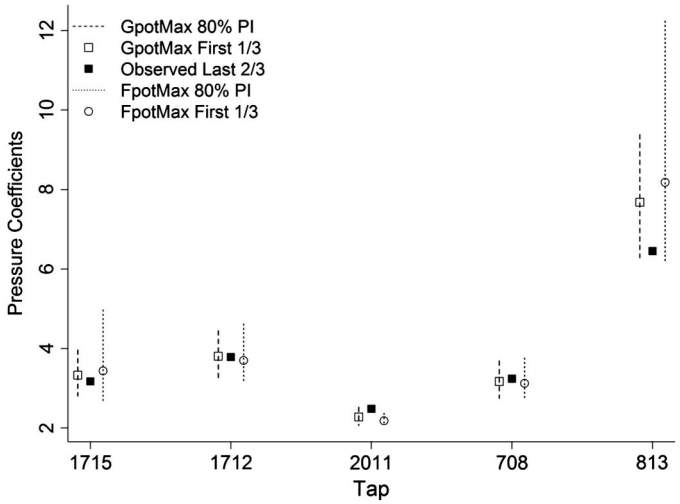

Fig. 12. Observed peaks and estimates by GpotMax and FpotMax of mean peaks and the 80% confidence intervals

Cross-Validation

To show that this approach works well in practice, it is not enough to show that it works well on the same data that were used to estimate the model parameters, as was the case for the results of Fig. 11. A good measure of a statistical method's ability is out-of-sample predictive accuracy. A common technique for estimating out-of-sample predictive accuracy is cross-validation. That is, the available data are partitioned into training and validation sets, the model parameters are estimated using the training set, and the trained model is then used to predict some property of the validation set. This exercise leveraged the series from five different taps, listed on the horizontal axis of Fig. 12. These five taps were chosen because they are representative of large pressure fluctuations, which commonly occur at roof corners and edges for negative pressures and at the middle of walls for positive pressures. A control with smaller pressure fluctuations (at the middle of the roof) also was included. Each series was partitioned into two pieces, one of 33 s and the other of 67 s. The shorter piece was used to estimate the model parameters and predict the overall peak of the longer piece.

Fig. 12 shows the results. The solid square depicts the observed peak from the 67-s partition; the open square shows the prediction of GpotMax for the 67-s piece, but trained on the 33-s piece; and the open circle illustrates the same results for FpotMax. The vertical dashed and dotted lines show 80% prediction intervals. These intervals are distinct from the confidence interval shown in Fig. 10. The difference between confidence intervals and prediction intervals is the difference between estimating a property of a population, such as the mean of the distribution of the peak value, and predicting a single outcome from a stochastic process, such as predicting the peak value from one new 67-s series.

The predictions from GpotMax and FpotMax shown in Fig. 12 are generally close to each other and to the observed peak. The largest discrepancy occurred for Tap 813, which is expected because of the larger pressure fluctuations at the roof corners compared with other locations. More importantly, Fig. 12 shows that the prediction intervals maintained their stated 80% coverage. It was expected that zero or one (of five) of the observed peaks would fall outside the prediction intervals. None of the observations fall outside the GpotMax bounds, and only one falls outside the FpotMax bounds, Tap 2011. This cross-validation exercise shows that the proposed approach performs well in practice.

Conclusions

This paper presents a method for estimating the distribution of the peak value of a time history of some length of interest, T_1, from an observed time history of length T, usually with $T < T_1$. The basis of the approach is a POT model, a two-dimensional Poisson process, first described by Pickands (1971). A reason for choosing a POT extreme value model is that it typically admits more data for estimation than do classic extreme value models used in more traditional epochal procedures. The approach proceeds by a series of seven steps: (1) ensure focus is on the distribution of a maximum; (2) choose whether to estimate the tail length parameter from the observed history or to set it to zero a priori; (3) decluster the observations to eliminate correlation between observations that are adjacent in time; (4) select an optimal threshold by balancing the tradeoff between variance and bias in estimation; (5) using the chosen threshold, estimate the model parameters by maximum likelihood, and calculate the Hessian matrix of the natural logarithm of the likelihood function at the maximum to be used subsequently in uncertainty quantification; (6) empirically construct the distribution of the peak by simulating many histories of length T_1 from the fitted model and saving the peak value of each simulated history; and (7) quantify uncertainty in the distribution of the peak by perturbing the parameter estimates according to the Hessian matrix calculated in Step 5 and the multivariate Gaussian distribution, and repeating Step 6 many times.

A major contribution of this paper to the existing body of literature is the estimation of the distribution of the peak value using a POT extreme value model. Epochal procedures obtain a closed form for the distribution of the peak value by leveraging classical extreme value models and order statistics. Such a closed form is not available for the two-dimensional Poisson process model employed here. This is because the number of threshold crossings is itself a random quantity. This difficulty is circumvented by using simulation to empirically construct the distribution of the peak value.

To illustrate the practical ability of this method, a cross-validation exercise was undertaken with histories from five example taps. These taps were selected from the available population of taps because they capture the largest fluctuations in wind pressure. The cross-validation exercise showed that the out-of-sample predictive accuracy of the method is quite good. Furthermore, and perhaps more importantly, it showed that the prediction intervals calculated according to the approximate bootstrap algorithm maintained their nominal coverage. This provides confidence that the statements of uncertainty are not gross overassessments or underassessments. The rigorous approach to quantifying uncertainty for the whole distribution of the peak is another major contribution of this paper to the current body of literature.

Finally, because the requisite computations for this approach are larger than for many currently popular procedures, fast and efficient software to accompany this paper has been developed and released as an *R* package. It is available at Pintar (2016) along with installation and use instructions.

Appendix. Tap Locations, Wind Directions, and Pressure Time Histories for Taps 1712, 2011, 708, and 813

Figs. 13–16 show the location of pressure taps 1712, 2011, 708, and 813, together with the wind direction and the time series used in this analysis.

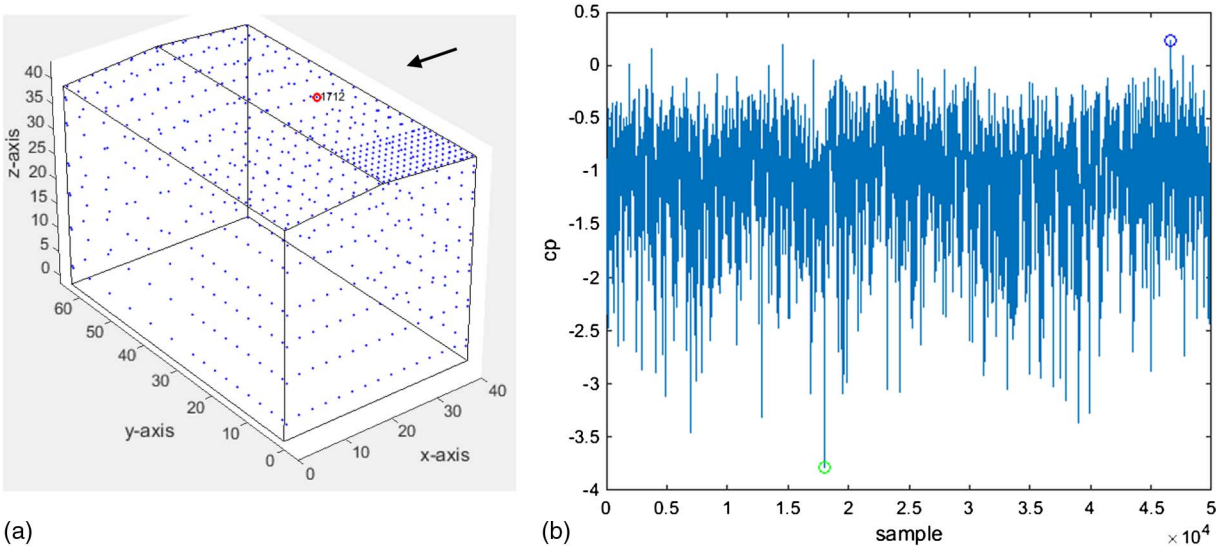

Fig. 13. (a) Location of Tap 1712; (b) time series of pressure for Tap 1712, wind 270° ($-x$), min = −3.7855

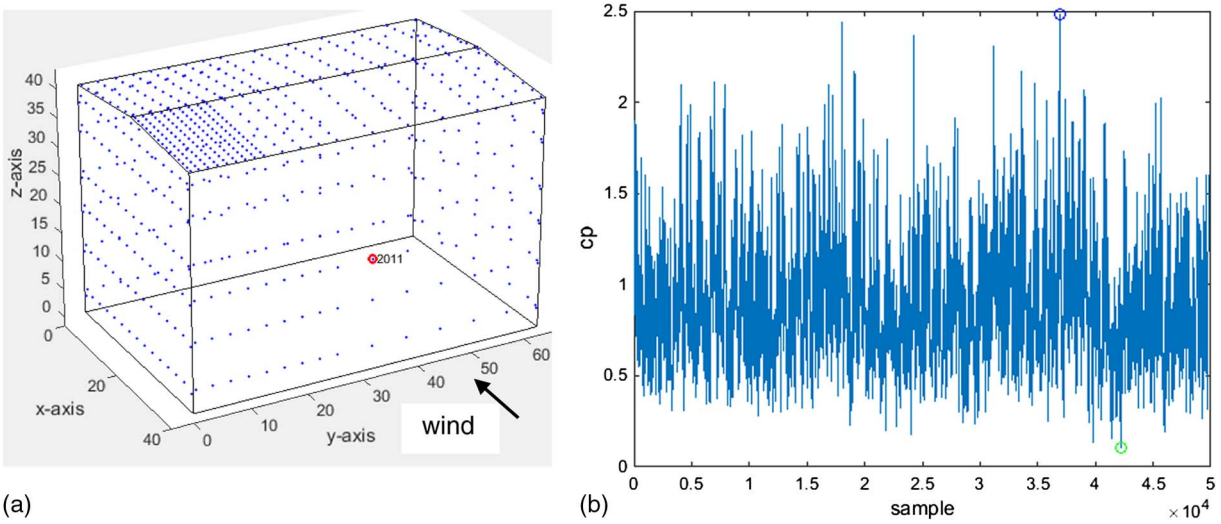

Fig. 14. (a) Location of Tap 2011; (b) time series of pressure for Tap 2011, wind 270° ($-x$), max = 2.4812

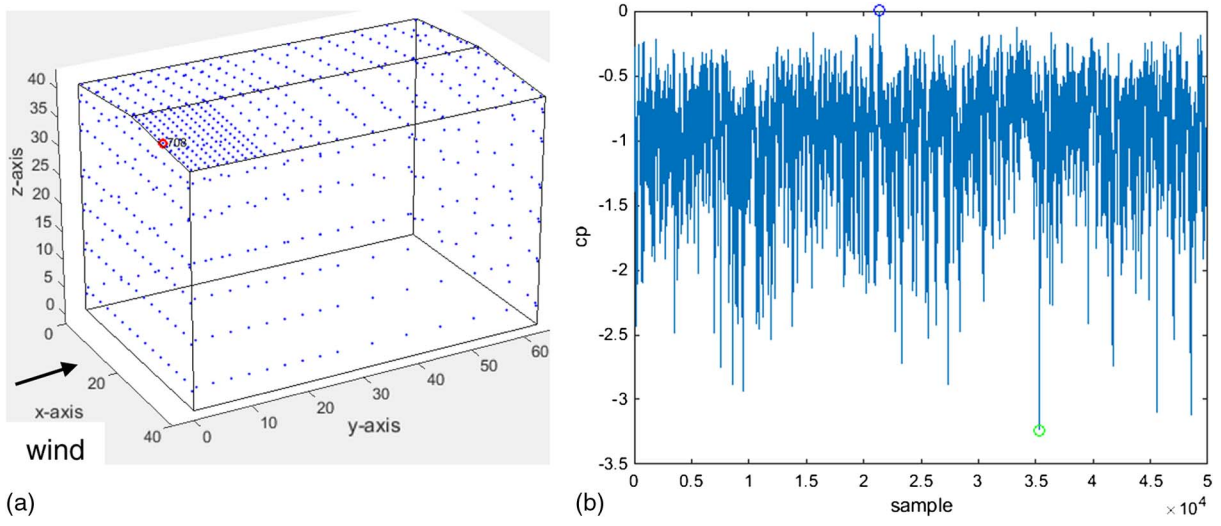

Fig. 15. (a) Location of Tap 708; (b) time series of pressure for Tap 708, wind 0° ($+y$), min = −3.2387

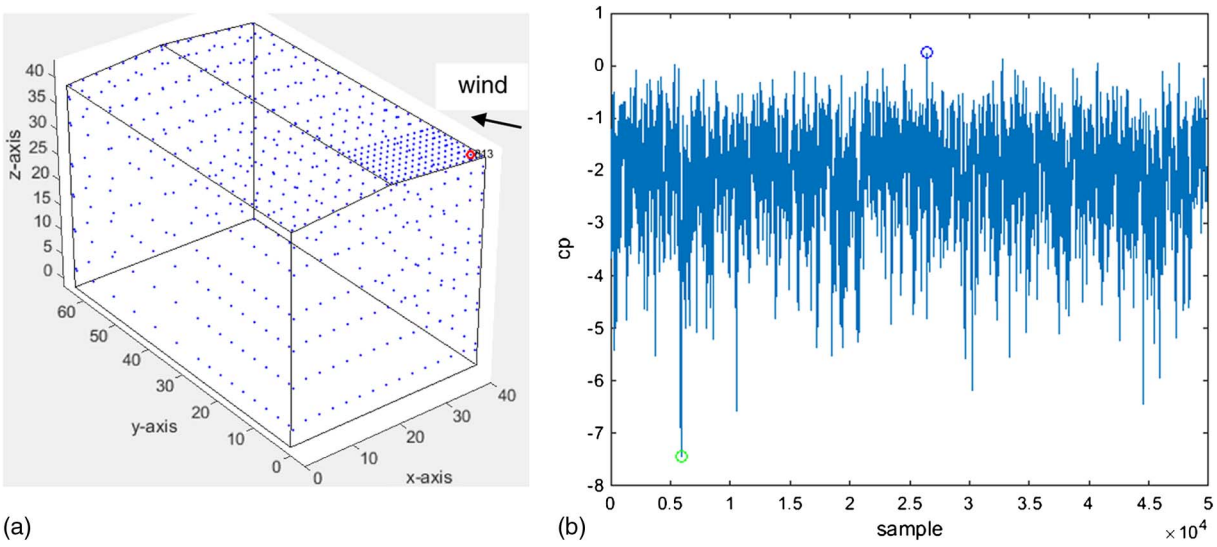

Fig. 16. (a) Location of Tap 813; (b) time series of pressure for Tap 813, wind 315°, min = −7.4482

References

Casella, G., and Berger, R. L. (2002). *Statistical inference*, Vol. 2, Duxbury, Pacific Grove, CA.

Coles, S. (2004). "The use and misuse of extreme value models in practice." Chapter 2, *Extreme values in finance, telecommunications, and the environment*, B. Finkenstädt and H. Rootzén, eds., Chapman & Hall/CRC, Boca Raton, FL, 79–100.

Cook, N. J. (1982). "Towards better estimation of extreme winds." *J. Wind Eng. Ind. Aerodyn.*, 9(3), 295–323.

Cook, N. J., and Mayne, J. R. (1979). "A novel working approach to the assessment of wind loads for equivalent static design." *J. Wind Eng. Ind. Aerodyn.*, 4(2), 149–164.

Cook, N. J., and Mayne, J. R. (1980). "A refined working approach to the assessment of wind loads for equivalent static design." *J. Wind Eng. Ind. Aerodyn.*, 6(1–2), 125–137.

Davenport, A. G. (1964). "Note on the distribution of the largest value of a random function with application to gust loading." *Proc. Inst. Civil Eng.*, 28(2), 187–196.

Efron, B., and Tibshirani, R. J. (1994). *An introduction to the bootstrap*, CRC Press, Boca Raton, FL.

Harris, R. I. (1999). "Improvement to the 'method of independent storms'." *J. Wind Eng. Ind. Aerodyn.*, 80(1–2), 1–30.

Harris, R. I. (2009). "XIMIS, a penultimate extreme value method suitable for all types of wind climate." *J. Wind Eng. Ind. Aerodyn.*, 97, 271–286.

Ho, T. C. E., Surry, D., Morrish, D., and Kopp, G. A. (2005). "The UWO contribution to the NIST aerodynamic database for wind loads on low buildings. Part 1: Archiving format and basic aerodynamic data." *J. Wind Eng. Ind. Aerodyn.*, 93(1), 1–30.

Lieblein, J. (1974). "Efficient methods of extreme value methodology." *NBSIR74-602*, National Bureau of Standards, Washington, DC.

Mannshardt-Shamseldin, E. C., Smith, R. L., Sain, S. R., Mearns, L. O., and Cooley, D. (2010). "Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data." *Ann. Appl. Stat.*, 4(1), 484–502.

NIST. (2004). "Extreme winds and wind effects on structures." ⟨http://www.itl.nist.gov/div898/winds/homepage.htm⟩ (Apr. 16, 2017).

Pasupathy, R. (2011). "Generating nonhomogeneous Poisson processes." *Wiley encyclopedia of operations research and management science*, Wiley, Hoboken, NJ.

Peterka, J. A. (1983). "Selection of local peak pressure coefficients for wind tunnel studies of buildings." *J. Wind Eng. Ind. Aerodyn.*, 13(1–3), 477–488.

Pickands III, J. (1971). "The two-dimensional Poisson process and extremal processes." *J. Appl. Probab.*, 8(04), 745–756.

Pickands III, J. (1994). "Bayes quantile estimation and threshold selection for the generalized Pareto family." *Proc., Conf. on Extreme Value Theory and Applications*, Kluwer Academic Publication, Boston.

Pintar, A. (2016). "potMax—Estimate the distribution of the maximum of a time series using peaks-over-threshold models." ⟨https://github.com/usnistgov/potMax⟩ (Aug. 16, 2017).

Pintar, A. L., Simiu, E., Lombardo, F. T., and Levitan, M. (2015). "Maps of non-hurricane non-tornadic winds speeds with specified mean recurrence intervals for the contiguous united states using a two-dimensional Poisson process extreme value model and local regression." ⟨http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.500-301.pdf⟩ (Aug. 16, 2017).

R [Computer software]. R Foundation for Statistical Computing, Vienna, Austria.

Resnick, S. I. (1992). *Adventures in stochastic processes*, Birkhäuser, Boston.

Rice, S. O. (1954). "Mathematical analysis of random noise." *Select papers on noise and stochastic processes*, N. Wax, ed., Dover, New York.

Sadek, F., and Simiu, E. (2002). "Peak non-Gaussian wind effects for database-assisted low-rise building design." *J. Eng. Mech.*, 10.1061/(ASCE)0733-9399(2002)128:5(530), 530–539.

Smith, R. L. (2004). "Statistics of extremes, with applications in environment, insurance, and finance." Chapter 1, *Extreme values in finance, telecommunications, and the environment*, B. Finkenstädt and H. Rootzén, eds., Chapman & Hall/CRC, Boca Raton, FL, 1–78.

Smith, R. L. (1989). "Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone." *Stat. Sci.*, 4(4), 367–377.